Integration geoConvention 2013 geoscience engineering partnership

Inexact Full Newton Method for Full Waveform Inversion using Simultaneous encoded sources

Amsalu Y. Anagaw, University of Alberta, Edmonton, Canada
aanagaw@ualberta.ca
and
Mauricio D. Sacchi, University of Alberta, Edmonton, Canada

Summary

In this paper, we present an inexact full Newton optimization method for the full waveform inversion algorithm in the frequency domain which utilizes simultaneous sources based upon the phase encoding technique. Tests show that the full Newton minimization method achieves a high convergence rate and a reasonably accurate reconstruction of the model parameters. Taking advantage of a direct solver based on LU decomposition, the full Newton minimization method can also be implemented in a matrix-free manner. Tests with this algorithm were conducted with the BP/EAGE velocity model and highlight its high performance capabilities.

Introduction

Seismic full waveform inversion's (FWI) main objective is to retrieve the Earth model that best describes the observed seismic data, which has lead to its inclusion to the arsenal of methods that determine subsurface velocity models (Tarantola, 1987; Pratt et al., 1998; Operto et al., 2007; Virieux and Operto, 2009; Hu et al., 2011). However, a couple of the main problems in FWI that limit its application is the computational cost of the inversion for multiple sources and receivers and its tendency to converge to local minima of its objective function. To solve the latter problem of converging to a local minima, a good initial velocity model or data with high-quality low frequency components are required. The computational cost of FWI can be reduced by utilizing simultaneous shooting techniques (Krebs et al., 2009; Ben-Hadj-Ali et al., 2011), since it is proportional to the number of sources within the experiment. The basic idea of the simultaneous shot method is to create super-shots through the summation of individual sources with a random encoding function (Romero et al., 2000). The drawback of the simultaneous source technique is that it introduces random cross-talk that arises from the correlation between shots. One way to reduce this cross-talk noise is by generating new random encoding super-shots in every iteration (Krebs et al., 2009).

FWI in the frequency domain is carried out in a sequential approach by selecting a few frequencies starting from low to higher frequencies (Sirgue and Pratt, 2004). The inversion result obtained from the first few frequencies (lower frequencies) is used to initialize the inversion for the next frequencies (higher frequencies) and so on. In order to reconstruct or invert for the model parameters from the measured seismic data, gradient based optimization methods such as steepest-descent and nonlinear conjugate gradients or Newtonian methods such as L-BFGS, Gauss-Newton and full Newton can be implemented. The Newtonian methods generally converge faster than the gradient based nonlinear conjugate gradient method, but this is at the expense of solving a denser system of linear equations (Hessian matrix) at each iteration which is computationally expensive. Of the Newtonian methods, the full Newton method is known to converge faster than the Gauss-Newton or L-BFGS methods. In this paper, a matrix-free full Newton method for FWI with simultaneous sources using a phase encoding technique is formulated. For each frequency in the proposed FWI algorithm, a new random encoding operator is generated. Numerical results on BP/EAGE velocity model (Billette and Brandsberg-Dahl, 2005) are presented and highlight the numerical efficiency of this method.

Theory

Full waveform inversion based on lease-squares objective function is the minimization of the l_2 norm of residual between the observed data $\mathbf{d}^{obs}$ and the model data $\mathbf{d}^{cal}$,

$$J(\mathbf{m}) = \frac{1}{2}\sum_{\omega}\sum_{s,r}^{N_s,N_r}(\mathbf{d}_{s,r}^{cal}(\omega)-\mathbf{d}_{s,r}^{obs}(\omega))^{\dagger}(\mathbf{d}_{s,r}^{cal}(\omega)-\mathbf{d}_{s,r}^{obs}(\omega)) + \mu R(\mathbf{m}), \quad (1)$$

where $\dagger$ is the complex conjugate transpose, $R(\mathbf{m})$ is the regularization term, μ is the regularization parameter and, N_s and N_r represent the number of sources and receivers respectively. For the sake of simplicity, the dependencies of spatial positions are not written explicitly. The minimization of the objective function, $J(\mathbf{m})$, is achieved through the Lagrangian constrained optimization method (Plessix, 2006; Akcelik et al., 2002)

$$\begin{aligned} &\underset{\mathbf{m}}{\text{minimize}} \quad J(\mathbf{m}) \\ &\text{subject to} \quad \mathbf{A}(\mathbf{m},\omega)\mathbf{p}_s(\omega) = \mathbf{S}(\omega)_s, \end{aligned} \quad (2)$$

where $\mathbf{A}(\mathbf{m},\omega)$ is the forward modelling operator, $\mathbf{p}_s(\omega)$ is the wavefield in space and $\mathbf{S}_s(\omega)$ is the encoded super-shots. Typically, equation [2] is solved using a direct solver based on an multifrontal LU decomposition of the finite-difference Helmholtz operator $\mathbf{A}$ into a lower and upper LU triangular decomposition scheme (Amestoy et al., 2001; Schenk and Gartner, 2004). This operator is quite sparse and, therefore, storable in memory. The main advantage of this method is that once the decomposition is performed and available for a given angular frequency ω and background velocity, the complex pressure field is efficiently solved for multiple sources using the forward and backward substitutions. The modelled data can then computed as $\mathbf{d}_{s,r}^{cal}(\omega) = \mathbf{r}\mathbf{p}_s(\omega)$, where $\mathbf{r}$ is a sampling operator. The encoded super-shots $\mathbf{S}(\omega)_s$ are obtained by

$$\mathbf{S}(\omega)_s = \Gamma\mathbf{D}_{\mathbf{R}}\mathbf{f}_s(\omega), \quad (3)$$

where $\mathbf{f}_s(\omega)$ is the monochromatic source term, Γ is the phase encoding function ($\Gamma(\phi) = e^{-i\phi}, \phi \in [0,2\pi]$) and $\mathbf{D}_R$ is the randomization operator that randomly picks the monochromatic sources. Using the Lagrangian multiplier, λ, the Lagrangian function for FWI with simultaneous sources becomes

$$\mathscr{L}(\mathbf{p},\mathbf{m},\lambda) = J(\mathbf{m}) + \sum_{\omega,s}^{N_w,N_s}\langle\lambda_s(\omega),\mathbf{A}(\mathbf{m},\omega)\mathbf{p}_s(\omega)-\mathbf{S}_s(\omega)\rangle_{\mathbf{x}} + \mu R(\mathbf{m}), \quad (4)$$

where $\langle,\rangle_{\mathbf{x}}$ is the scalar product in $\mathbf{x}$. From the first-order optimality conditions and setting $\nabla\mathscr{L}_{\mathbf{p},\mathbf{m},\lambda} = \mathbf{0}$, also known as the Karush-Kuhn-Tucker (KKT) conditions, the reduced gradient, $\mathbf{g}$, of the objective function becomes

$$\mathbf{g} = ((\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p}_s(\omega))\lambda_s^{\dagger}(\omega) + \mu\nabla_{\mathbf{m}}R. \quad (5)$$

From the second-order optimality conditions, the reduced Newton of the objective function becomes

$$\begin{aligned} \mathbf{H} =& ((\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s})^*(\mathbf{A}^*)^{-1}\mathbf{r}^T\mathbf{r}\mathbf{A}^{-1}(\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s} + \mu\nabla_{\mathbf{m}}^2 R \\ &+ \mathbf{K} - \mathbf{B}^*\mathbf{A}^{-1}\mathbf{C} - \mathbf{C}^*(\mathbf{A}^*)^{-1}\mathbf{B}, \end{aligned} \quad (6)$$

where

$$\begin{aligned} \mathbf{K} &= (\nabla_{\mathbf{m}}^{\mathbf{2}}\mathbf{A})\mathbf{p}_s(\omega)\lambda_s^{\dagger}(\omega) \\ \mathbf{B} &= (\nabla_{\mathbf{m}}\mathbf{A})^*\lambda_s(\omega) \\ \mathbf{C} &= (\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p}_s(\omega) \\ \lambda_s(\omega) &= (\mathbf{A}^*)^{-1}\mathbf{r}^T(\mathbf{r}\mathbf{p}^{cal}(\omega) - \mathbf{d}^{obs}(\omega)). \end{aligned} \quad (7)$$

Note that λ_s is the backpropagation of the residual wavefields at the receiver positions.
In Newton's formulation, the model perturbation update is computed by solving the following linear system of equations

$$\mathbf{H}\triangle\mathbf{m} = -\mathbf{g}. \tag{8}$$

The Hessian, $\mathbf{H}$, is a dense and full matrix and, in large scale problems, it is often too expensive to store or solve using direct solvers. In practice, the model perturbation is computed through iterative techniques. Here, we adopt the conjugate gradient least-squares (CGLS) method to solve equation [8]. The CGLS scheme requires only the action of the Hessian on the model perturbation; matrix vector product, which can be implemented in matrix-fee manner. For a single frequency, once the matrix $\mathbf{A}$ is factorized using LU decomposition, a set of solutions for multiple shots can be achieved by forward and backward substitutions at a relatively low computational time. Therefore, for each frequency the matrix $\mathbf{A}$ is factorized only once and the action of the full Hessian $\mathbf{H}$ on the model perturbation $\mathbf{m}$ is achieved on the fly at each CGLS iteration by the solutions of four forward problems. In the Newton method, we adopt the early termination of the CGLS iteration and update the model parameters using a line search method.

Below is the pseudo-code of the inexact full Newton method. Notations: ω_g :- data weight, nw :- number of frequencies within a group frequency nwg. This is only applicable for simultaneous multi-frequency inversion.

Pseudo code: Multiscale inexact full$-$Newton method
Start $\leftarrow \mathbf{m}_0$
do igr = 1, ngw $\leftarrow$*o*ver group frequency
Start $\leftarrow \mathbf{m}_k$
$\Gamma, D_R \leftarrow$*g*enerate random operators
do $k = 1, max\_iter$
Compute the gradient :
$\nabla_{\mathbf{m}} J(\mathbf{m}_k) = \sum_{iw}^{nw} \omega_g(iw) \nabla_{\mathbf{m}} J(\mathbf{m}_k)_{iw}$
$\longrightarrow$ *solve two forward problems (FP)*
$\triangle\mathbf{m}_k = -\left[\sum_{iw}^{nw} \Re e(\mathbf{H})_{iw}\right]^{-1} \nabla_{\mathbf{m}} J(\mathbf{m}_k)$
$\longrightarrow$ *solve using CGLS*
$\longrightarrow$ *solve four FP for each CGLS itet.,* ω
Update: $\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha\triangle\mathbf{m}_k$
$\longrightarrow$ α *:- step length*
enddo
enddo

Examples

The BP/EAGE velocity model is used to test our FWI algorithm. The original BP/EAGE velocity model is 67km long and 12km deep and built on a 6.25m x 6.25m grid size. This velocity model was then re-gridded onto a 100m x 100m spatial grid. Figure 1 [a] depicts the true BP/EAGE velocity model and Figure 1 [b] shows the initial linearly increasing velocity model (from 1400m/s to 4500m/s) used for FWI. The synthetic data was generated with a total number of 225 sources and 338 receivers. For the FWI, a set of nine discrete frequencies were selected from approximately 0.2Hz - 5Hz. In order to recover the long wavelength components of the velocity model and mitigate the local minima effects of FWI, one has to start the inversion from very low frequency data. The low frequency data components for this velocity model inversion have also been used by Hu et al. (2011).

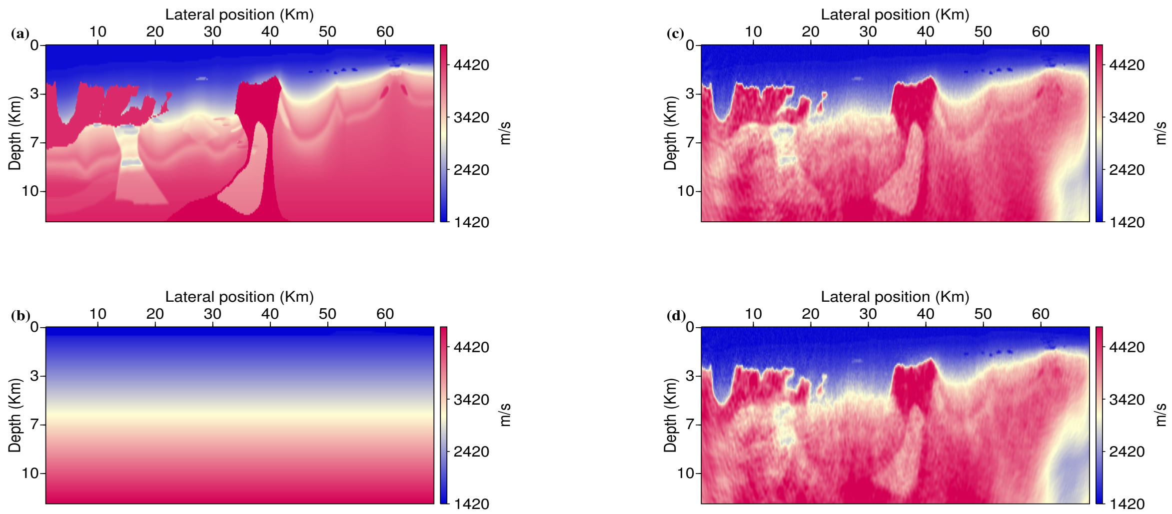

Figure 1 *BP/EAGE velocity model (a), linearly increasing velocity model with depth used as starting model for inversion (b). (c) and (d) are reconstructed velocity model using simultaneous sources with 15 and 9 super-shots respectively.*

For each frequency, a maximum of 20 full Newton iterations were computed. Each conjugate gradient least-squares was computed with 15 iterations. Figure 1c & d are the reconstructed velocity models using a total of 15 and 9 super-shots where each super-shot is constructed by randomly encoding 15 and 25 monochromatic sources, respectively. Both numerical inversions reproduce a model that is quite comparable to the original velocity model. Most notably, both the salt body structures and shallow anomalies are properly reconstructed. For each inversion frequency, a new random source and phase encoding operator was generated. These figure images demonstrate our full Newton FWI algorithm provides very competitive results in terms of resolution and quality of the inverted model.

Conclusion

We present an inexact full Newton full waveform inversion algorithm that uses the simultaneous sources technique. The full waveform inversion problem is based on a quadratic matrix-free full Newton method, where the objective function converges aquatically. By utilizing the advantage of a direct solver based on the LU decomposition of the forward modelling operator, for each frequency, the operator matrix is factorized only once and the action of the Hessian on the model perturbation is computed on the fly with the CGLS algorithm. Through tests with the BP/EAGE velocity model, we demonstrate that the full Newton FWI algorithm can recover an accurate and high resolution image of the velocity model. In order to speed up the computational time, we adapted the simultaneous source encoding techniques. The simultaneous source technique reduces the computational cost proportional to the number of super-shots per every iteration, which in turn reduces the over all cost drastically, which becomes even more important when dealing with costly three-dimension numerical simulations.

Acknowledgements

The authors are grateful to the sponsors of Signal Analysis and Imaging Group (SAIG) at the University of Alberta.

References

Akcelik, V., G. Biros, and O. Ghattas, 2002, Parallel multiscale gauss-newton-krylov methods for inverse wave propagation: 41.

Amestoy, P., I. Duff, J. L'Excellent, and J. Koster, 2001, A fully asynchronous multifrontal solver using distributed dynamic scheduling: SIAM JOURNAL ON MATRIX ANALYSIS AND APPLICATIONS, **23**, 15–41.

Ben-Hadj-Ali, H., S. Operto, and J. Virieux, 2011, An efficient frequency-domain full waveform inversion method using simultaneous encoded sources: Geophysics, **76**, R109–R124.

Billette, F. and S. Brandsberg-Dahl, 2005, The 2004 BP velocity benchmark., *in* 67th Annual Internat. Mtg., EAGE, Expanded Abstracts, B035. EAGE.

Hu, W., A. Abubakar, T. M. Habashy, and J. Liu, 2011, Preconditioned non-linear conjugate gradient method for frequency domain full-waveform seismic inversion: Geophysical Prospecting, **59**, 477–491.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, **74**, WCC177–WCC188.

Operto, S., J. Virieux, P. Amestoy, J.-Y. L'Excellent, L. Giraud, and H. B. H. Ali, 2007, 3d finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study: Geophysics, **72**, SM195–SM211.

Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

Pratt, G., C. Shin, and Hicks, 1998, Gauss−newton and full newton methods in frequency−space seismic waveform inversion: Geophysical Journal International, **133**, 341–362.

Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436.

Schenk, O. and K. Gartner, 2004, Solving unsymmetric sparse systems of linear equations with PARDISO: Future Generation Computer Systems, **20**, 475–487. International Conference on Computational Science (ICCS 2002), Amsterdam, Netherlands, APR 21-24, 2002.

Sirgue, L. and R. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: GEOPHYSICS, **69**, 231–248.

Tarantola, A., 1987, Inverse problem theory. methods for data fitting and model parameter estimation: Amsterdam: Elsevier.

Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.